PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Gallien L, Thornhill AH, Zurell D, Miller JT, Richardson DM. 2019 Global predictors of alien plant establishment success: combining niche and trait proxies. *Proc. R. Soc. B* **286**: 20182477.
http://dx.doi.org/10.1098/rspb.2018.2477

Received: 2 November 2018
Accepted: 10 December 2018

Subject Category:
Global change and conservation

Subject Areas:
ecology, plant science, environmental science

Keywords:
anthropogenic disturbance, biological invasions, ecological niche, historical biogeography, open-access database

Author for correspondence:
Laure Gallien
e-mail: laure.gallien@univ-grenoble-alpes.fr

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4391327.

THE ROYAL SOCIETY PUBLISHING

Global predictors of alien plant establishment success: combining niche and trait proxies

Laure Gallien[1,2], Andrew H. Thornhill[3,4], Damaris Zurell[5], Joseph T. Miller[4,6] and David M. Richardson[1]

[1]Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Matieland 7602, South Africa
[2]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LECA (Laboratoire d'Ecologie Alpine), 38000 Grenoble, France
[3]Australian Tropical Herbarium, James Cook University, Cairns, Queensland 4870, Australia
[4]Centre for Australian National Biodiversity Research, National Research Collections, Black Mountain, Canberra, Australian Capital Territory, Australia
[5]Geography Department, Humboldt-Universität zu Berlin, Berlin 10099, Germany
[6]Office of International Science and Engineering, National Science Foundation, Alexandria, VA 22314, USA

LG, 0000-0003-4882-1580; DZ, 0000-0002-4628-3558

Biological invasions are on the rise globally. To reduce future invasions, it is imperative to determine the naturalization potential of species. Until now, screening approaches have relied largely on species-specific functional feature data. Such information is, however, time-consuming and expensive to collect, thwarting the screening of large numbers of potential invaders. We propose to resolve such data limitations by developing indicators of establishment success of alien species that can be readily derived from open-access databases. These indicators describe key features of successfully established aliens, including estimates of potential range size, niche overlap with human-disturbed environments, and proxies of species traits related to their palaeoinvasions and local dominance capacities. We demonstrate the utility of this new approach by applying it to two large and highly invasive plant groups: Australian acacias and eucalypts. Our results show that these indicators robustly predict establishment successes and failures in each clade independently, and that they can cross-predict establishment in these two clades. Interestingly, the indicator identified as most important was species potential range size on Earth, a variable too rarely considered as a predictor. By successfully identifying key features that predispose Australian plants to naturalize, we provide an objective and cost-effective protocol for flagging high-risk introductions.

1. Background

The threat posed by invasive species on biodiversity and ecosystem functioning is increasing rapidly worldwide [1]. Surprisingly, although research on biological invasions is soaring, it remains difficult to anticipate which species will establish in a new region and, thus, to design cost-effective prevention plans. The difficulty is largely the result of two conceptual and methodological challenges. First, for many plant groups, the ecological features that favour establishment success are poorly understood [2]. Second, although key correlates of establishment success have been identified for certain species, these rely on measuring relevant functional traits, which is impractical to do for a large number of species. For example, factors, such as resistance of species to disturbance or capacity for rapid adaptation, are difficult and time-consuming to evaluate [3]. Therefore, we are still lacking widely applicable indicators for accurately flagging potential future invaders.

Indicators of alien establishment success (also called naturalization) can be ascertained based on a simple screening approach using widely accessible data on occurrence of species, climate and phylogeny. In particular, we derive four potential indicators of species naturalization from these data, related to species niche and trait characteristics. They include species' environmental niche characteristics (niche breadth and niche availability on Earth), niche overlap with significantly human-disturbed environments, and proxies of traits related to their establishment capacities (number of 'palaeoinvasions' and level of local dominance).

In this study, we test the predictive power of these four naturalization indicators for two Australian clades, in a region that has served as a factory for invasive woody plants, for reasons that are not yet well understood despite extensive investigations. These two clades are Australian 'acacias' (*Acacia sensu stricto*; *ca* 1020 species, including 386 species known to have been introduced to new regions, 71 of which have established self-sustaining populations; or 18% naturalization success) and 'eucalypts' (a clade of three genera: *Angophora*, *Corymbia*, and *Eucalyptus*; *ca* 830 species, of which 82 out of 374 known introductions have resulted in naturalization, or 22% naturalization success) [4–6]. We focus on naturalized species: those that were able to establish self-sustaining populations. These species have not necessarily been reported to show rapid and problematic spread in the new areas or to outcompete native flora, but could do so in the future.

We aimed to model naturalization potential using four key establishment indicators. First, we expect the likelihood of naturalization to increase when the species can establish and flourish across a wide range of abiotic and biotic conditions in their new ranges. Consequently, we calculated *niche breadth* as an indicator of the range of climatic conditions in which introduced species could survive, and the *potential global range size* as an indicator for the probability of a species to find a suitable introduction site on Earth [7]. Second, we expect the likelihood of naturalization to increase when a species is heavily used (e.g. for ornamental horticulture, forestry) and transported (intentionally or accidentally) by humans. We thus used species *niche overlap with human-disturbed environments* as an indicator for the probability of human-mediated transport and cultivation. This overlap measure has the additional benefit of accounting for the fact that humans tend to settle in non-stressful environments, which generally corresponds to fast-growing phenotypes in plants, a frequent characteristic of successful invaders [8]. These three niche-related factors are not completely independent. For example, large niche breadth may correlate with large potential global range size and high overlap with human-disturbed environments. Nevertheless, we regard them as complementary indicators useful for disentangling specific cases, such as species with narrow niche breadth but large potential geographical extent on Earth. Third, we derived *the number of palaeoinvasions* (i.e. the number of biogeographic regions that ancestors of a species have successfully colonized over evolutionary timescales) as an indicator of species capacities to colonize novel biogeographical regions [9]. Biogeographical movements of species are often favoured by traits involved in long-distance dispersal, resistance to disturbance and rapid adaptation [3], but these traits are difficult to measure, especially for long-lived organisms such as trees. We thus used an indirect approach introduced by Gallien *et al.* [9], which is based on the assumption that the number of 'palaeoinvasions' [10] undertaken by the ancestors of an extant species (prior to human intervention) is indicative of its current capacity to establish into novel regions. Finally, we calculated a measure of species spatial aggregation (the fractal dimension of their area of occupancy (AOO)) that has been shown to correlate with their *capacity for local dominance* [11]. Sparse and fragmented distributions tend to reflect local extinction processes, whereas denser and more aggregated distributions tend to reflect species local dominance [11,12]. All these indicators were derived from global open-access databases containing information on native range distributions of each species (Australasian Virtual Herbarium; https://avh.chah.org.au/), species-level phylogenies for acacias [13] and eucalypts [6] (as used to infer the number of palaeoinvasions, see method section), global climate (WorldClim v.2.0) and human-disturbed environments (Human Footprint database).

2. Methods

(a) Study species

Many Australian plants, including hundreds of species of acacias and eucalypts, have been moved around the world over the past 150 years for supporting and provisioning ecosystem services such as soil erosion control, timber, perfume or tannin production [14–16]. The *Acacia* clade includes 1020 species, almost all endemic to Australia, of which at least 315 are known to have been planted outside Australia but have not been reported as naturalized, 48 are naturalized (reproducing consistently) and 23 are invasive (spreading from sites of introduction) outside of Australia (i.e. 71 successful versus 315 failed establishments, or 18% naturalization success [4,17]). The eucalypt clade includes 830 species, most of them endemic to Australia, of which 292 are known to have been planted outside Australia but are not reported as naturalized, 74 are naturalized and eight are invasive outside of Australia (i.e. 82 successful versus 292 failed establishments, or 22% naturalization success [17]). It should be noted that all species identified as introduced (leading to either successful or failed naturalization) were introduced with roughly the same intensity (species introduced only to botanical gardens were not considered). Categorization of introduction status for species follows the definitions of Blackburn *et al*. [18].

(b) Primary data compilation

Species native range was estimated from occurrence records in the Australasian Virtual Herbarium (avh.chah.org.au; accessed in July 2016; *ca* 244 000 records for acacias and *ca* 248 000 records for eucalypts; sub-species were aggregated to the species level). To ensure maximal comparability between the acacia and eucalypt datasets, the data source and methods for cleaning and organizing the data followed protocols as set out in Richardson *et al.* [4] and Robertson *et al.* [19], and the minimal resolution accepted was 5 km. We manually removed records of species from well outside the known native range in Australia (several species from eastern Australia have been widely planted in Western Australia and vice versa [4]). The data editing and cleaning resulted in *ca* 171 000 records for 1020 acacia species (154 occurrences on average per species) and 170 000 records of 742 eucalypt species (202 occurrences on average per species).

Environmental conditions were obtained from the global climatic database WorldClim v.2.0 (http://www.worldclim.org).

We extracted the 19 bioclimatic layers at a resolution of 2.5 arc min (to match the resolution of the occurrence dataset). Because many bioclimatic variables are highly correlated (as derived from monthly temperature and rainfall values), we selected the four variables that revealed the lowest correlation coefficients on the Australian continent (pairwise correlation less than 0.6): mean diurnal temperature range (bio2), maximum temperature of the warmest month (bio5), annual precipitation (bio12) and precipitation seasonality (bio15).

The phylogeny of acacias was generated as in Mishler *et al*. [13] and includes 639 Australian species (63% of the total species list). The alignment was analysed using maximum-likelihood (ML) using the CIPRES RAxML Blackbox utility. The resultant ML phylogeny was dated using penalized likelihood in r8s [20]. Calibrations for each group were obtained from previous molecular dating studies and are as follows. The root of Australian acacias was calibrated using the 95% age range estimate for the divergence between Australian acacias and their sister groups 23.0–34.06 Ma (same age estimates as obtained by Miller *et al*. [21]). Other calibrations were obtained from the 95% age range estimates from the same study [21] and placed on the following acacia groups: *mulga* (9.26–11.14 Ma); *plurinerves* (11.17–13.47 Ma); *mulga* + *plurinerves* (12.07–15.59 Ma); *botrycephela* (11.06–13.62 Ma); *victoriae* (6.16–7.44 Ma); *pull* (16.21–19.46 Ma); *murraya* (10.02–12.02 Ma); *tetragonophylla* (7.57–15.73 Ma); and *dunnii* (9.41–17.98 Ma). The phylogeny of eucalypts was obtained from [6] and includes 711 Australian eucalypt species (86% of the total species list). Eucalypt calibration points were selected from previous studies [22,23]. Two fossil calibrations were used: a *Eucalyptus* macrofossil 51.69–52.13 Ma [24], and an *Angophora* + *Corymbia* fossil pollen 45–47.0 Ma [25]; as well as one secondary calibration of the eucalypts (all three genera) crown 51.2–54.7 Ma [22].

(c) Ecological niche characteristics

The ecological niches of species were quantified and compared in Australian environmental space. We did this by first extracting the climatic conditions over the whole of Australia and running a principal component analysis (PCA). The PCA allowed us to find the main axes of climatic differences across different regions in the continent, and to obtain the environmental space in which to contrast different species niches. For each species, we then located areas in the PCA space where it was recorded as present, and used a 95% alpha-hull around the presences as an estimate of their ecological niche. Although this is a simplistic estimation of species ecological niche, it allows us to study all species in the same environmental space, and to compute relatively quickly the niche dimensions for the hundreds of species targeted in our screening protocol. We then characterized these niche estimates in terms of niche breadth and niche availability on Earth (hereafter called *potential global range size*). The climatic niche breadth for each species was computed as the total area covered by its alpha-hull in the PCA space. The potential global range size of each species (i.e. the frequency of occurrence of its suitable climate on Earth) was estimated by simply counting the number of 2.5 arc min pixels on Earth (excluding Australia) that fell within species alpha-hull climatic niche. It should be noted that estimating species ecological niche based solely on their distribution in their native range may underestimate their actual niche breadth, as well as their capacities to adapt to new environmental conditions [26,27]. However, this is a necessary constraint when predictions need to be made on species that have never been introduced (e.g. for building blacklists of species that pose a risk of becoming invasive).

(d) Niche overlap with human-disturbed environments

Human-disturbed environments were identified based on the open-access human footprint (HF) database (SEDAC database [28]). HF provides an index of anthropogenic impacts on the environment [29] and indicates the degree of disturbance, ranging from 0 (pristine) to 100 (highly disturbed) based on multiple indicators including human population pressure, human land use and infrastructure and human access [29]. HF was obtained at 1 km resolution and aggregated to match the resolution of the occurrence and climatic data. Sites were considered to be significantly disturbed by humans if the HF index was equal to or greater than 30 (the results were robust to the choice threshold; see electronic supplementary material, appendix S1). The overlap between species climatic niche and the climate in human-disturbed environments was estimated using Schoener's D metric (corrected for differences in relative availability of environments in the R package *ecospat* [30]). D varies between 0 (no overlap) and 1 (complete overlap). It should be noted that this variable is a synthetic estimate, which combines the environmental suitability at sites of introduction, local disturbance regimes and species detectability.

(e) Number of palaeoinvasions

Species capacity for colonizing novel biogeographical regions was estimated following the recommendations of Gallien *et al*. [9], as the number of biogeographic regions that their ancestors have successfully colonized over evolutionary timescales. To estimate how the ancestors of extant species have moved during their evolutionary history, we followed three steps: (i) we identified the main biogeographical regions of acacias and eucalypts in Australia, (ii) we reconstructed lineage colonization history and (iii) we counted the number of colonized regions per lineage. First, the biogeographical regions were defined using BIODIVERSE v.1.99_007 following the methodology proposed by González-Orozco *et al*. [31,32]. The definition of biogeographical regions can be based on either taxonomic or phylogenetic turnover measures. To integrate the uncertainty associated with the choice of turnover measure, we quantified both, which produced two sets of biogeographical regions per clade (two sets for acacias, and two for eucalypts; see electronic supplementary material, appendix S2). For each of these sets of biogeographical regions, all acacia and eucalypt species were labelled as being native to one or more regions (based on the overlap between their presences and the regions). Second, we estimated the historical biogeography of clades using the R package BioGeoBEARS [33], following the procedure used by Gallien *et al*. [9] (the inputs being acacia and eucalypt phylogenies and biogeographical regions). In brief, we ran the three most commonly used models (i.e. DIVA, DEC and BayArea), as well as their more complex versions incorporating founder-event speciation (hereafter DIVA + J, DEC + J and BayArea + J). For each model, we allowed ancestral distributions to extend to all available regions, with the specific constraint that before the aridification period (about 10 to 20 Ma), the central arid region was absent. We combined all model predictions using AIC-weighted model-averaging to obtain single-ensemble estimates [9]. Because the date of the aridification period remains uncertain, we repeated this analysis with two different starting times of the arid period (10 Ma and 20 Ma). Third, based on the biogeographical movement reconstruction, we were able to trace back in time how many regions were colonized by each lineage. This was done by counting the number of colonized regions between the root of the phylogeny to each tip. This overall procedure led us to obtain four different estimates of numbers of past colonization events per extant species. These four estimates per species denote uncertainties inherent to this approach (i.e. identification of biogeographical region from taxonomic versus phylogenetic turnover, and dating of the Australian aridification period 10 Ma versus 20 Ma), and were used to track these uncertainties in the following analyses. It

can be noted that uncertainties in phylogenetic dating could similarly be incorporated to the analyses (e.g. when fossils are sparse).

(f) Local dominance capacity indicators

Species dominance capacity was estimated with the fractal dimension of the AOO by species following the recommendations of Wilson *et al.* [11]. In brief, for each species, we estimated its AOO at two spatial scales. We first summed the areas of occupied 5-km-squares, then aggregated these to 25 km-squares and again summed the areas of the occupied 25-km-squares. Log_{10}AOO at each scale was plotted against log_{10} (side of grid square (in kilometre)). The fractal dimension (D_{ij}) was the slope of this scale–area curve, subtracted from 2 [34]. A maximum value of $D_{ij} = 2$ indicates that occupied fine-scale cells completely fill each occupied coarse-scale cell (most aggregated distributions), while a minimum value of $D_{ij} = 0$ indicates that each occupied fine-scale cell is located in a separate coarse-scale cell (sparsest distributions).

(g) Describing species naturalization potential

For each clade, we estimated models that relate naturalization success to our four key establishment indicators: species ecological niche characteristics, niche overlap with human-disturbed environments, number of palaeoinvasions (as a proxy for traits related to the colonization of novel environments) and a proxy for their local dominance capacities. Naturalization was considered successful for all introduced species that had established self-perpetuating populations, and failed for species that are known to have been introduced but which have not established self-perpetuating populations. The acacias and eucalypts account for 386 and 374 introduced species, respectively. However, constraints on data availability (including their inclusion in a phylogeny) reduced these numbers to 259 acacias (56 naturalized versus 203 non-naturalized) and 317 eucalypts (73 naturalized versus 244 non-naturalized) for model building. Specifically, for each clade, we predicted species naturalization success (with a binomial logit model where successfully naturalized species were coded as 1 and failed ones as 0) as a function of species (i) niche characteristics (niche breadth and potential global range size), (ii) niche overlap with human-disturbed environments, (iii) number of palaeoinvasions and (iv) local dominance capacities (as estimated by the level of spatial aggregation of their populations). Correlation between these predictor variables was below $r = 0.6$ except for the correlation between potential global range size and niche breadth ($r = 0.85$ and $p < 0.001$ for acacias; $r = 0.82$ and $p < 0.001$ for eucalypts). Despite the strong correlation in the latter two variables, we chose to keep all the variables in the initial models to test for their combined effect. The number of variables in the model was reduced by performing a variable selection with a forward–backward step-wise procedure based on the AIC criterion. To account for the non-independence of the species due to shared ancestry, we used phylogenetic regressions (R package *phylolm* [35]). We also accounted for the uncertainties in our estimates of the number of palaeoinvasions by repeating the overall modelling procedure with our four estimates of the number of ancestral colonizations (cf. Number of palaeoinvasion section), which left us with four models per clade. We then estimated the variable importance in each model using a permutation accuracy importance algorithm, as suggested by Strobl *et al.* [36] (using 100 permutations). The principle was to apply a random permutation of each variable to mimic the absence of the variable in the model. Importance is then estimated according to the difference in prediction accuracy with and without the permutation.

(h) Predicting species establishment

The predictive performance of each model was evaluated in terms of their discrimination and classification ability using the area under the relative operating characteristic curve (AUC), the proportion of predictions correctly classified (PCC), the rate of under-prediction (UPR) and the rate of over-prediction (OPR). The AUC ranges from 0.5 (prediction not better than random) to 1 (perfect prediction). A model with an AUC higher than 0.8 is usually classified as relatively good [37,38]. UPR and OPR measures of false prediction rates in the models follow indices presented in [39].

First, we assessed within-clade predictive performance by running a cross-validation procedure via repeated sample splitting which randomly splits the data into a calibration group (80%) and a validation group (20%) of species. We then separately fitted the models to the calibration subset of species and predicted naturalization success for the holdout species (validation dataset). Prediction accuracy for these holdout species was assessed using AUC and PCC. The splitting procedure was repeated 50 times. Second, for evaluating the between-clade extrapolation power of our screening approach, we used the model calibrated on 100% of the acacia data (including only the predictors selected with the AIC-based variable selection) to predict naturalization success in eucalypts, and *vice versa*. Between-clade prediction accuracy of the acacia and eucalypt models was assessed using AUC and PCC.

All statistical analyses were performed with the software R v.3.4.0 [40].

3. Results and discussion

For each clade, we estimated phylogenetic logistic regression models to relate species establishment success versus failure to the proposed indicators. AIC-based variable selection and randomization procedures were used to evaluate the relative importance of the different indicator groups (the predictors selected for each clade are presented in figure 1). For both acacias and eucalypts, the most important predictor of establishment success, by far, was the potential global range size (with 61% and 76% variable importance for acacias and eucalypts, respectively; figure 1). This confirms that the alien species that successfully establish are those that have larger potential global range sizes compared to species that fail to establish. Interestingly, even though potential global range size of species is positively correlated with their niche breadth ($r = 0.85$ and $p < 0.001$ for acacias; $r = 0.82$ and $p < 0.001$ for eucalypts), the latter variable was never retained in our models. This indicates that potential global range size is a better predictor of establishment success than niche breadth. Successful aliens are thus not necessarily those species with the broadest climatic niches as traditionally assumed [8,41], but rather those that can inhabit the largest ranges. Indeed, having a large niche breadth does not guarantee that suitable climates for the taxon are well represented on Earth [41] (e.g. *Acacia pellita* or *Eucalyptus pulchella*). Additionally, in both clades we found establishment success to increase significantly where climatic niche overlapped with the climate of human-disturbed sites (including urban areas, cultivated areas and areas with dense road networks) although this effect was more pronounced in acacias (with 22% variable importance for acacias and 3% for eucalypts). This result highlights that species with the best chance of naturalizing are those that can establish populations in areas where humans either

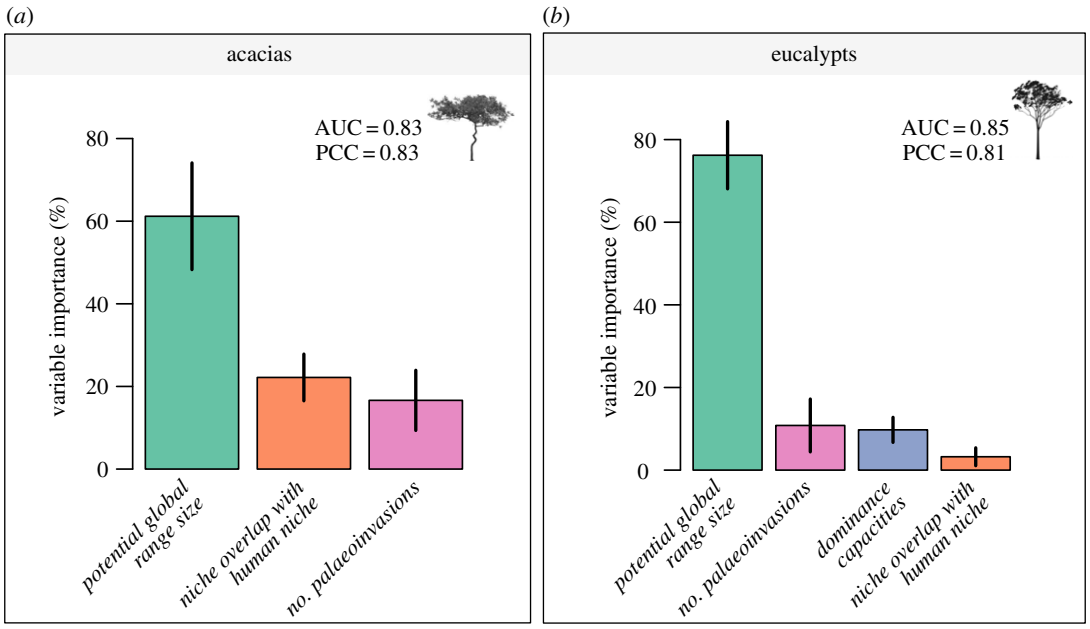

Figure 1. Relative importance of species naturalization predictors for (*a*) acacia and (*b*) eucalypt species. For each clade, binomial phylogenetic regressions were used to discriminate between successful versus failed naturalizations. An AIC-based selection was applied to select only significant predictors, and a randomization procedure to quantify the relative importance of predictors. The coloured bars represent the median variable importance across all models (four models per clade were estimated to quantify the uncertainty associated with the measures of palaeoinvasions), and the vertical lines indicate the standard deviation of the results around these medians. The model performance was evaluated with AUC scores (ranging from 0.5: prediction not better than random, to 1: perfect prediction; AUC $\geq$ 0.8 usually classified as good), and the proportion of predictions correctly classified (PCC).

live, cultivate plants or travel; such species do not necessarily have the largest niche size. In line with our expectations, establishment success systematically increased with the *number of palaeoinvasions* (variable importance 17% and 11% for acacias and eucalypts, respectively). An effect of species' *capacity for local dominance* on establishment success was only apparent in eucalypts (variable importance 10%).

Based on these phylogenetic regressions, we assessed the ability of our indicators to predict the likelihood of other species becoming naturalized. This within-clade cross-validation (80–20% random split-sample approach) showed good predictive power for both acacias (median AUC = 0.81, PCC = 0.84, UPR = 0.21, OPR = 0.22) and eucalypts (median AUC = 0.81, PCC = 0.82, UPR = 0.22, OPR = 0.17). Between-clade prediction accuracy (from acacia model to eucalypt data, and *vice versa*) showed similarly good performances, with only a marginal decrease in AUC of 2% and in PCC of 3–5% compared to within-clade cross-predictions (figure 2). These results show that the likelihood of a species naturalizing following introduction to new regions can be broadly assessed using simple indicators derived from widely accessible databases. Importantly, these indicators are general enough to cross-predict between distantly related clades (i.e. models calibrated on one clade can provide useful predictions onto another) while remaining within an acceptable range of error [42]. Although these indicators are generic, the models can be re-calibrated for clades originating from other biogeographical regions to accommodate particular features of the focal taxa. Overall, our approach provides a novel risk assessment methodology that can supplement current approaches that are often time-consuming and/or require expert knowledge (e.g. the UK risk assessment protocol takes on average 19 h per species, which would translate into more than 5 years of work for our 576 species [43,44]).

Banning the international movement of all species is certainly the safest prevention approach in the context of highly invasive clades such as acacias and eucalypts ('ban them all' strategy [15]). However, this option is often unrealistic in taxa with large commercial or other value. An alternative is to develop defendable 'black lists' of species. Black lists are based on the quantification of invasion risk of species to provide robust and quantifiable justification for the regulation of species that pose a high risk of becoming invasive. Although such lists are widely advocated as a practical approach for simplifying the process of regulating intentional introductions [45], objective protocols for populating such lists are still lacking, even for groups with many invasive species such as acacias and eucalypts (more subjective protocols based on expert knowledge have been developed [46] but have shown limited efficacy and reproducibility [42,47]). Our screening approach provides exactly such a probabilistic assessment of species invasiveness (table 1), and thus has the potential for being widely adopted. It is noteworthy that each country can select its own 'banning threshold' for compiling a black list according to transparent decisions relating to the cost–benefit balance in misclassification (i.e. the ratio of the costs of banning a species that is harmless and potentially useful, and the cost of introducing a future invader [48]).

Although only utilizing a small number of crude predictors, our models had predictive accuracies of 84% and 82% for the acacia and eucalypt clades, respectively. Nonetheless, it is important to emphasize that this approach is useful as a first screening for large numbers of species; it is unlikely to produce flawless predictions of which species are sure to become naturalized. Indeed, a few species were predicted to be naturalized while there is, as yet, no evidence of naturalization (e.g. *Corymbia tessellaris*, *Eucalyptus pauciflora*). This may reflect that key variables are still missing in the models and/or an invasion debt, meaning that the species have not

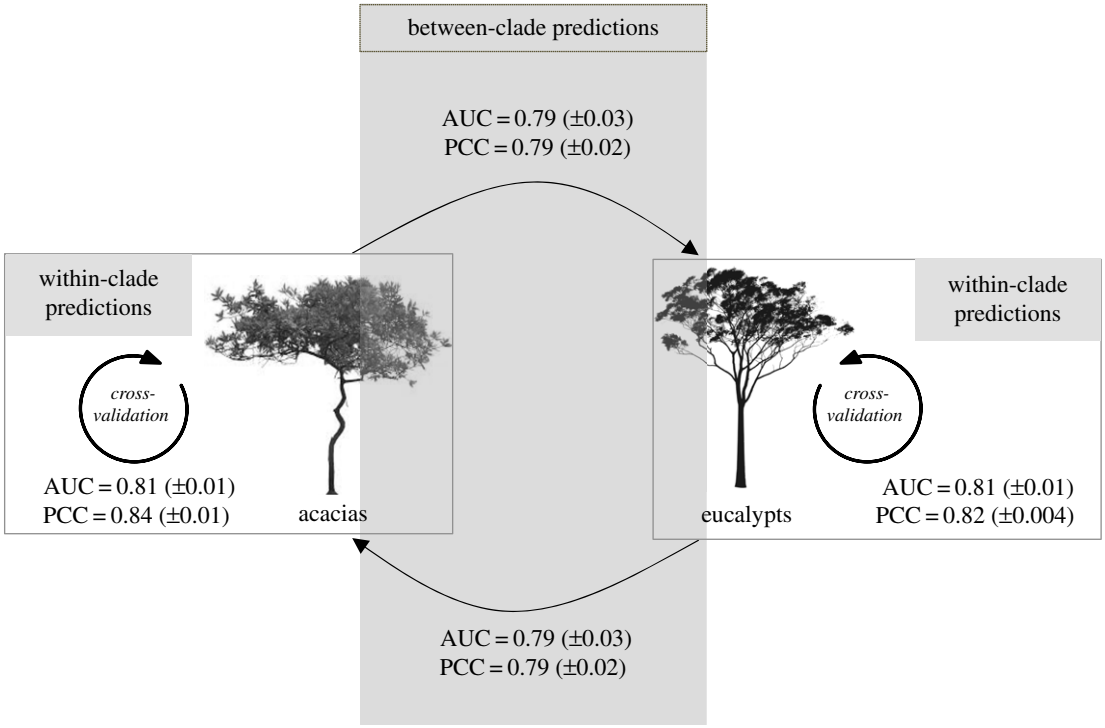

Figure 2. Performance of the screening approach. For each clade (acacias and eucalypts), we used binomial phylogenetic regressions to identify the main indicators of naturalization success, which we evaluated both within and between clades. Within-clade performance was assessed using a split-sample cross-validation (80% of the data used for calibration and 20% for validation, repeated 100 times). We assessed between-clade performance by calibrating the models on 100% of the species of one clade and predicting to 100% of the species of the other clade. The model performance was evaluated with AUC scores (ranging from 0.5: prediction not better than random, to 1: perfect prediction; AUC $\geq$ 0.8 usually classified as good), and the proportion of predictions correctly classified (PCC).

Table 1. A proposed black list of the 10 non-introduced acacias and eucalypts that are predicted to have the highest risk of naturalization if introduced.

acacias	eucalypts
Acacia rostellifera	*Eucalyptus leptophylla*
Acacia siculiformis	*Corymbia clarksoniana*
Acacia linifolia	*Corymbia erythrophloia*
Acacia applanata	*Eucalyptus lockyeri*
Acacia gilbertii	*Eucalyptus platyphylla*
Acacia multisiliqua	*Eucalyptus phenax*
Acacia continua	*Eucalyptus chartaboma*
Acacia dorothea	*Eucalyptus conspicua*
Acacia bidentata	*Eucalyptus barberi*
Acacia alpina	*Eucalyptus vegrandis*

had enough time or opportunities to establish in adventive ranges but may do so in the future [49]. Some other species were predicted as being unlikely to naturalize but have been reported as naturalized in at least one region of the world (e.g. *Acacia mountfordiae*, *E. diversicolor*). These cases are particularly interesting, as most of these species have relatively small native range sizes (with native range sizes 2.15 and 4.33 times smaller than the successfully predicted invaders for acacias and eucalypts, respectively), which could have biased the estimation of their environmental niche (e.g. underestimation of species tolerance [27,50,51], or capacity to adapt to novel conditions [26]). Such misclassification of species with small native ranges may be overcome in the future by building more complete estimates of environmental niches of species (and of its evolution), for instance by integrating experimental measures. However, both our within- and between-clade validations underline that even such simple estimates of native niche of species are useful for explaining and predicting naturalization success outside their native ranges.

Our proposed screening approach based on simple indicators is a global-scale evaluation of the risk of naturalization, and provides a single synthetic measure per species. Further refinements could easily be implemented for particular study regions and study species. For instance, if the aim is to provide a black list of species for a specific region, our models could be coupled with spatially explicit predictions from climatic niche models to evaluate which of the high-risk species would find large suitable areas in natural habitats and should thus be prohibited. Additionally, if a given study region lacks detailed information on its native species (e.g. if data on native ranges are only available at very coarse spatial scales), then characteristics of other species can be sought to complement the models (e.g. the altitudinal range over which the species occurs, or measures of its tolerance of frost, heat and shade).

The model could also be extended to predict the transition from naturalization to invasion. For this, additional data on species functional characteristics would be required—for instance including measures of species competitive superiority, fecundity and plasticity [3,52–55]—and could simply be added to the model. Although such traits are often not freely available for a large set of species, the emergence of large collaborative trait databases (e.g. TRY; https://www.try-db.org/TryWeb/Home.php) may soon make such additions to screening models feasible.

In conclusion, our results show that naturalized species from two Australian clades share key biogeographic and evolutionary features: large global potential range sizes, many instances of palaeoinvasions, and strong overlap with human-disturbed environments. The identification of these features through our screening approach and simple indicators opens new opportunities for the prediction of species naturalization potential in other plant taxonomic groups. Moreover, because our indicators are derived from open-access global databases, they are easily and rapidly applicable to large numbers of species, thereby filling a critical gap in the design of global black lists of alien species.

Data accessibility. All data used in this study were downloaded from the open-access databases listed in the Methods.

Authors' contribution. L.G. and D.M.R. conceived the study; L.G., A.H.T., J.T.M. and D.M.R. collected and compiled the datasets; L.G. performed all analyses with advice from A.H.T. and D.Z.; L.G. wrote the first draft of the manuscript that was then edited by all authors. All authors gave final approval for publication.

Competing interests. We have no competing interests.

Funding. Funding was provided by the DST-NRF Centre for Invasion Biology and the National Research Foundation (grant 85417 to D.M.R.). D.Z. acknowledges support from the German Science Foundation (DFG grant no. ZU 361/1-1). This manuscript includes work done by J.T.M. while serving at the National Science Foundation. The views expressed in this paper do not necessarily reflect those of the National Science Foundation or the United States Government.

Acknowledgements. We thank F. C. Boucher for many interesting discussions during the course of the study.

References

1. Vilà M, Hulme PE. 2017 *Impact of biological invasions on ecosystem services*. Cham: Springer International Publishing
2. Gallien L, Carboni M. 2017 The community ecology of invasive species: where are we and what's next? *Ecography* (*Cop.*) **40**, 335–352. (doi:10.1111/ecog.02446)
3. Theoharides KA, Dukes JS. 2007 Plant invasion across space and time: factors affecting nonindigenous species success during four stages of invasion. *New Phytol.* **176**, 256–273. (doi:10.1111/j.1469-8137.2007.02207.x)
4. Richardson DM, Carruthers J, Hui C, Impson FAC, Miller JT, Robertson MP, Rouget M, Le Roux JJ, Wilson JRU. 2011 Human-mediated introductions of Australian acacias: a global experiment in biogeography. *Divers. Distrib.* **17**, 771–787. (doi:10.1111/j.1472-4642.2011.00824.x)
5. Booth TH. 2014 Modern tree colonisers from Australia into the rest of the world. In *Invasion biology and ecological theory* (eds HHT Prins, IJ Gordon), pp. 304–323. New York, NY: Cambridge University Press.
6. González-Orozco C *et al.* 2016 Phylogenetic approaches reveal biodiversity threats under climate change. *Nat. Clim. Chang.* **6**, 1110–1114. (doi:10.1038/nclimate3126)
7. Higgins SI, Richardson DM. 2014 Invasive plants have broader physiological niches. *Proc. Natl Acad. Sci. USA* **111**, 10 610–10 614. (doi:10.1073/pnas.1406075111)
8. Castro-Díez P, Godoy O, Saldaña A, Richardson DM. 2011 Predicting invasiveness of Australian acacias on the basis of their native climatic affinities, life history traits and human use. *Divers. Distrib.* **17**, 934–945. (doi:10.1111/j.1472-4642.2011.00778.x)
9. Gallien L, Saladin B, Boucher FC, Richardson DM, Zimmermann NE. 2016 Does the legacy of historical biogeography shape current invasiveness in pines? *New Phytol.* **209**, 1096–1105. (doi:10.1111/nph.13700)
10. Richardson DM, Pyšek P, Rejmánek M, Barbour MG, Panetta FD, West CJ. 2000 Naturalization and invasion of alien plants: concepts and definitions. *Divers. Distrib.* **6**, 93–107. (doi:10.1046/j.1472-4642.2000.00083.x)
11. Wilson JW, Thomas CD, Fox R, Roy DB, Kunin WE. 2004 Spatial patterns in species distribution reveal biodiversity change. *Nature* **432**, 393–396. (doi:10.1038/nature03036.1.)
12. Hui C, Richardson DM, Robertson MP, Wilson JRU, Yates CJ. 2011 Macroecology meets invasion ecology: linking the native distributions of Australian acacias to invasiveness. *Divers. Distrib.* **17**, 872–883. (doi:10.1111/j.1472-4642.2011.00804.x)
13. Mishler BD, Knerr N, González-Orozco CE, Thornhill AH, Laffan SW, Miller JT. 2014 Phylogenetic measures of biodiversity and neo-and paleo-endemism in Australian *Acacia*. *Nat. Commun.* **5**, 4473. (doi:10.1038/ncomms5473)
14. Reichard SH, Hamilton C. 1997 Predicting invasions of woody plants introduced into North America. *Conserv. Biol.* **11**, 193–203. (doi:10.1046/j.1523-1739.1997.95473.x)
15. Wilson JRU *et al.* 2011 Risk assessment, eradication, and biological control: global efforts to limit Australian *Acacia* invasions. *Divers. Distrib.* **17**, 1030–1046. (doi:10.1111/j.1472-4642.2011.00815.x)
16. Miller JT, Hui C, Thornhill AH, Gallien L, Le Roux JJ, Richardson DM. 2017 Is invasion success of Australian trees mediated by their native biogeography, phylogenetic history, or both? *AoB Plants* **9**, plw80. (doi:10.1093/aobpla/plw080)
17. Rejmánek M, Richardson DM. 2013 Trees and shrubs as invasive alien species—2013 update of the global database. *Divers. Distrib.* **19**, 1093–1094. (doi:10.1111/ddi.12075)
18. Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JRU, Richardson DM. 2011 A proposed unified framework for biological invasions. *Trends Ecol. Evol.* **26**, 333–339. (doi:10.1016/j.tree.2011.03.023)
19. Robertson MP, Visser V, Hui C. 2016 Biogeo: an R package for assessing and improving data quality of occurrence record datasets. *Ecography* (*Cop.*) **39**, 394–401. (doi:10.1111/ecog.02118)
20. Sanderson MJ. 2002 Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* **19**, 101–109. (doi:10.1093/oxfordjournals.molbev.a003974)
21. Miller JT, Murphy DJ, Ho SYW, Cantrill DJ, Seigler D. 2013 Comparative dating of *Acacia*: combining fossils and multiple phylogenies to infer ages of clades with poor fossil records. *Aust. J. Bot.* **61**, 436. (doi:10.1071/BT13149)
22. Thornhill AH, Popple LW, Carter RJ, Ho SYW, Crisp MD. 2012 Are pollen fossils useful for calibrating relaxed molecular clock dating of phylogenies? A comparative study using Myrtaceae. *Mol. Phylogenet. Evol.* **63**, 15–27. (doi:10.1016/j.ympev.2011.12.003)
23. Thornhill AH, Ho SYW, Külheim C, Crisp MD. 2015 Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. *Mol. Phylogenet. Evol.* **93**, 29–43. (doi:10.1016/j.ympev.2015.07.007)
24. Gandolfo MA, Hermsen EJ, Zamaloa MC, Nixon KC, González CC, Wilf P, Cúneo NR, Johnson KR. 2011 Oldest known *Eucalyptus* macrofossils are from South America. *PLoS ONE* **6**, e21084. (doi:10.1371/journal.pone.0021084)
25. Thornhill AH, Macphail MK. 2012 Fossil myrtaceous pollen as evidence for the evolutionary history of the Myrtaceae: a review of fossil *Myrtaceidites* species. *Rev. Palaeobot. Palynol.* **176–177**, 1–23. (doi:10.1016/j.revpalbo.2012.03.003)
26. Gallien L, Thuiller W, Fort N, Boleda M, Alberto FJ, Rioux D, Lainé J, Lavergne S. 2016 Is there any evidence for rapid, genetically-based, climatic niche expansion in the invasive common ragweed? *PLoS ONE* **11**, e0152867. (doi:10.1371/journal.pone.0152867)
27. Gallien L, Münkemüller T, Albert CH, Boulangeat I, Thuiller W. 2010 Predicting potential distributions of invasive species: where to go from here? *Divers. Distrib.* **16**, 331–342. (doi:10.1111/j.1472-4642.2010.00652.x)
28. Sanderson EW, Jaiteh M, Levy MA, Redford KH, Wannebo AV, Woolmer G. 2002 The human footprint and the last of the wild. *Bioscience* **52**, 891–904. (doi:10.1641/0006-3568(2002)052[0891:THFATL]2.0.CO;2)

29. Woolmer G, Trombulak SC, Ray JC, Doran PJ, Anderson MG, Baldwin RF, Morgan A, Sanderson EW. 2008 Rescaling the human footprint: a tool for conservation planning at an ecoregional scale. *Landsc. Urban Plan.* **87**, 42–53. (doi:10.1016/j.landurbplan.2008.04.005)
30. Di Cola V *et al*. 2017 ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography* (*Cop.*). **40**, 774–787. (doi:10.1111/ecog.02671)
31. González-Orozco CE, Laffan SW, Knerr N, Miller JT. 2013 A biogeographical regionalization of Australian *Acacia* species. *J. Biogeogr.* **40**, 2156–2166. (doi:10.1111/jbi.12153)
32. Laffan SW, Lubarsky E, Rosauer DF. 2010 Biodiverse, a tool for the spatial analysis of biological and related diversity. *Ecography* (*Cop.*) **33**, 643–647. (doi:10.1111/j.1600-0587.2010.06237.x)
33. Matzke NJ. 2013 Probabilistic historical biogeography: new models for founder event speciation, imperfect detection, and fossils allow improved accuracy and model testing. *Front. Biogeogr.* **5**, 242–248. (doi:10.21425/F55419694)
34. Kunin WE. 1998 Extrapolating species abundance across spatial scales. *Science* **281**, 1513–1515. (doi:10.1126/science.281.5382.1513)
35. Ho L, Ané C. 2014 A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* **63**, 397–408. (doi:10.1093/sysbio/syu005)
36. Strobl C, Boulesteix A-L, Zeileis A, Hothorn T. 2007 Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinf.* **8**, 25. (doi:10.1186/1471-2105-8-25)
37. Swets KA. 1988 Measuring the accuracy of diagnostic systems. *Science* **240**, 1285–1293. (doi:10.1126/science.3287615)
38. Araújo MB, Pearson RG, Thuiller W, Erhard M. 2005 Validation of species–climate impact models under climate change. *Glob. Chang. Biol.* **11**, 1504–1513. (doi:10.1111/j.1365-2486.2005.01000.x)
39. Barbosa MA, Real R, Muñoz AR, Brown JA. 2013 New measures for assessing model equilibrium and prediction mismatch in species distribution models. *Divers. Distrib.* **19**, 1333–1338. (doi:10.1111/ddi.12100)
40. R Development Core Team. 2017 *R: a language and environment for statistical computing*. https://www.R-project.org/.
41. Peterson AT, Soberon J, Pearson RG, Anderson RP, Martinez-Meyer E, Nakamura M, Araujo MB. 2011 *Ecological niches and geographic distributions*. Princeton, NJ: Princeton University Press.
42. Gordon DR, Lieurance D, Flory SL. 2017 Predicted versus actual invasiveness of climbing vines in Florida. *Biol. Invasions* **19**, 2375–2384. (doi:10.1007/s10530-017-1448-7)
43. Kumschick S, Richardson DM. 2013 Species-based risk assessments for biological invasions: advances and challenges. *Divers. Distrib.* **19**, 1095–1105. (doi:10.1111/ddi.12110)
44. Verbrugge LN., Leuven RSEW, van de Velde G. 2010 Evaluation of international risk assessment protocols for exotic species. Department of Environmental Science, Report. Institute for Water and Wetland Research, Radboud University Nijmegen, The Netherlands.
45. Burgiel S, Perrault AM. 2011 Black, white, and gray lists. In *Encyclopedia of biological invasions* (ed. M Simberloff, D, Rejmánek), pp. 75–77. Berkeley, CA: University of California Press.
46. Daehler CC, Denslow JS, Ansari S, Kuo HC. 2004 A risk-assessment system for screening out invasive pest plants from Hawaii and other Pacific Islands. *Conserv. Biol.* **18**, 1–9. (doi:10.1111/j.1523-1739.2004.00066.x)
47. Essl F, Nehring S, Klingenstein F, Milasowszky N, Nowack C, Rabitsch W. 2011 Review of risk assessment systems of IAS in Europe and introducing the German–Austrian Black List Information System (GABLIS). *J. Nat. Conserv.* **19**, 339–350. (doi:10.1016/J.JNC.2011.08.005)
48. Smith CS, Lonsdale WM, Fortune J. 1999 When to ignore advice: invasion predictions and decision theory. *Biol. Invasions* **1**, 89–96. (doi:10.1023/A:1010091918466)
49. Rouget M, Robertson MP, Wilson JRU, Hui C, Essl F, Renteria JL, Richardson DM. 2015 Invasion debt—quantifying future biological invasions. *Divers. Distrib.* **22**, 445–456. (doi:10.1111/ddi.12408)
50. Gallien L, Douzet R, Pratte S, Zimmermann NE, Thuiller W. 2012 Invasive species distribution models—how violating the equilibrium assumption can create new insights. *Glob. Ecol. Biogeogr.* **21**, 1126–1136. (doi:10.1111/j.1466-8238.2012.00768.x)
51. Atwater DZ, Ervine C, Barney JN. 2018 Climatic niche shifts are common in introduced plants. *Nat. Ecol. Evol.* **2**, 34–43. (doi:10.1038/s41559-017-0396-z)
52. Van Kleunen M, Dawson W, Schlaepfer D, Jeschke JM, Fischer M. 2010 Are invaders different? A conceptual framework of comparative approaches for assessing determinants of invasiveness. *Ecol. Lett.* **13**, 947–958. (doi:10.1111/j.1461-0248.2010.01503.x)
53. Rejmánek M, Richardson DM. 1996 What attributes make some plant species more invasive? *Ecology* **77**, 1655–1661. (doi:10.2307/2265768)
54. Allen WL, Street SE, Capellini I. 2017 Fast life history traits promote invasion success in amphibians and reptiles. *Ecol. Lett.* **20**, 222–230. (doi:10.1111/ele.12728)
55. Capellini I, Baker J, Allen WL, Street SE, Venditti C. 2015 The role of life history traits in mammalian invasion success. *Ecol. Lett.* **18**, 1099–1107. (doi:10.1111/ele.12493)